УДК: 004.8; 004.2 https://doi.org/10.15407/jai2025.04.153

Д. І. Симонов[1], **І. О. Деменко**[2]
[1,2]Інститут кібернетики ім. В. М. Глушкова НАН України, Україна
пр. Академіка Глушкова, 40, м. Київ, 03187
[1]denys.symonov@gmail.com
[2]ivandemenkofifa@gmail.com
[1]https://orcid.org/0000-0002-6648-4736
[2]https://orcid.org/0009-0008-2683-6523

ВИКОРИСТАННЯ МУЛЬТИАГЕНТНИХ ПІДХОДІВ ДЛЯ ВАЛІДАЦІЇ МЕДИЧНИХ АНАЛІТИЧНИХ СИСТЕМ

D. Symonov[1], **I. Demenko**[2]
[1,2]V. M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Ukraine
Academician Glushkov Avenue, 40, Kyiv, 03187
[1]denys.symonov@gmail.com
[2]ivandemenkofifa@gmail.com
[1]https://orcid.org/0000-0002-6648-4736
[2]https://orcid.org/0009-0008-2683-6523

USING MULTI-AGENT APPROACHES FOR THE VALIDATION OF MEDICAL ANALYTICAL SYSTEMS

Анотація. Сучасні медичні аналітичні системи, що використовують технології штучного інтелекту, потребують підвищеної достовірності, стабільності та відтворюваності результатів. Традиційні методи тестування не враховують складну структуру гібридних архітектур, у яких поєднуються модулі машинного навчання, експертні системи та статистичні механізми. У статті запропоновано мультиагентну концепцію автоматизованої валідації медичних аналітичних систем, у якій процес перевірки реалізується через взаємодію автономних агентів – Generator, Validator, Evaluator, Data та Coordinator. Використання генеративних моделей як активних агентів дозволяє створювати синтетичні клінічні сценарії, що розширюють тестовий простір і забезпечують оцінювання систем у рідкісних або критичних умовах. Розроблено математичну модель, яка описує інформаційну динаміку агентів, а також аналітично доведено збіжність системи до інформаційної рівноваги. Експериментальна верифікація показала, що запропонований підхід зменшує ентропійну невизначеність у три рази, підвищує стабільність рішень на 20–25 % та скорочує час перевірки у 15 разів. Отримані результати підтверджують ефективність мультиагентної валідації як основи для створення адаптивних, самонавчальних і прозорих систем контролю якості медичних ШІ-рішень.

Ключові слова: мультиагентні системи, валідація, штучний інтелект, генеративні моделі, медичні аналітичні системи, тестування, інтелектуальні агенти, цифрова медицина.

Abstract. Modern medical analytical systems powered by artificial intelligence (AI) require enhanced reliability, stability, and reproducibility of analytical outcomes. Conventional validation techniques are insufficient for complex hybrid architectures that combine machine learning algorithms, expert-based reasoning, and statistical inference modules. This paper proposes a multi-agent concept for automated validation of medical analytical systems, in which the validation process is realized through the interaction of autonomous agents – Generator, Validator, Evaluator, Data, and Coordinator. The use of generative models as active agents enables the creation of synthetic clinical scenarios, expanding the testing space and allowing evaluation of systems under rare or critical conditions. A mathematical model of the proposed multi-agent dynamics is developed to describe the information exchange and coordination mechanisms among agents, and the convergence of the system toward informational equilibrium is analytically demonstrated. Experimental verification confirms that the proposed approach reduces entropic uncertainty by a factor of three, increases decision stability by 20–25%, and decreases validation time by more than 15 times compared to traditional methods. The obtained results demonstrate the effectiveness of the multi-agent validation framework as a foundation for developing adaptive, self-learning, and transparent quality-control environments for AI-based medical systems.

Keywords: multi-agent systems, validation, artificial intelligence, generative models, medical analytical systems, testing, intelligent agents, digital medicine.

Вступ

Сучасна медицина переживає фазу глибокої цифрової трансформації, у центрі якої перебувають технології штучного інтелекту (ШІ). Вони вже сьогодні застосовуються для інтерпретації медичних зображень, прогнозування ризиків захворювань, стратифікації пацієнтів і

підтримки прийняття клінічних рішень. Алгоритми машинного навчання та глибинні нейронні мережі дозволяють виявляти складні нелінійні закономірності у великих масивах даних, забезпечуючи продуктивність, недосяжну для традиційних статистичних методів. Водночас зростає кількість публікацій, які фіксують, що висока точність моделі не гарантує її клінічну надійність: навіть незначні варіації у структурі вхідних даних можуть призвести до істотних змін у прогнозах. Така нестійкість пов'язана із гетерогенністю джерел даних, відмінностями у протоколах збору, стохастичною природою алгоритмів навчання та відсутністю формалізованих критеріїв валідації для гібридних ШІ-систем. Унаслідок цього виникає ризик аналітичного розсинхрону – ситуації, коли локально правильні висновки не узгоджуються на системному рівні, що потенційно загрожує хибними клінічними рішеннями. Забезпечення достовірності, стійкості та відтворюваності результатів стає, отже, не лише технічною, а й етичною вимогою сучасної медицини. У цьому контексті завдання валідації медичних аналітичних систем, тобто підтвердження їх коректності, надійності та логічної узгодженості, набуває фундаментального значення для розвитку цифрової медицини та впровадження ШІ у клінічну практику.

Постановка проблеми

Ключовою науковою проблемою є відсутність формалізованого, динамічного механізму узгодження результатів валідації у складних медичних аналітичних системах, де алгоритми машинного навчання, експертні модулі та статистичні компоненти функціонують у різних часових і структурних масштабах. Така гетерогенність призводить до появи суперечливих висновків між підсистемами, втрати стабільності рішень і неможливості забезпечити цілісну оцінку достовірності. Традиційні підходи базуються на статичних тестових наборах і не враховують взаємозалежність процесів генерації даних, перевірки й оцінювання, що унеможливлює адаптацію системи до змін клінічного контексту. Відсутність координаційного механізму між етапами валідації породжує інформаційний дисбаланс: результати окремих модулів не інтегруються в єдину логічну систему перевірки. Саме цю проблему пропонується розв'язати за допомогою мультиагентного підходу, у межах якого автономні агенти – генератори, валідатори, оцінювачі та координатори – взаємодіють у замкненому циклі, колективно забезпечуючи когерентність, узгодженість і стійкість процесу валідації. Така взаємодія формує саморегульоване середовище, здатне еволюційно підтримувати рівновагу та підвищувати надійність медичних ШІ-систем.

Мета дослідження

Метою дослідження є розроблення інтегрованої мультиагентної моделі автоматизованої валідації медичних аналітичних систем, у якій процеси генерації, перевірки, оцінювання та координації реалізуються через взаємодію спеціалізованих агентів у єдиному інформаційному циклі. Запропонований підхід спрямований на формування динамічного механізму узгодження результатів валідації, здатного забезпечувати когерентність дій між агентами та еволюційне самоналаштування системи відповідно до змін структури даних або клінічного контексту. Генеративні моделі в межах цієї концепції виступають інструментом побудови синтетичних сценаріїв і тестових кейсів, що дозволяє розширити простір перевірки без порушення етичних чи конфіденційних вимог. Досягнення поставленої мети забезпечить створення самоорганізованого середовища валідації, у якому точність, стабільність і відтворюваність медичних ШІ-висновків підтримуються через багаторівневу взаємодію агентів, що узгоджують свої дії за інформаційними та ентропійними критеріями.

Аналіз останніх досліджень і публікацій

Проблема достовірності та стабільності роботи систем штучного інтелекту в медицині активно досліджується у світовій науковій

спільноті. У працях Тополя Е., Естеви А. та Раззака М. [1–3] розглядаються питання використання машинного навчання для діагностики та прогнозування захворювань. Автори відзначають, що моделі ШІ можуть бути чутливими до зміни даних, мати упередження або втрачати точність у нових клінічних сценаріях, що підкреслює потребу у формалізованих методах валідації перед практичним застосуванням.

Більшість існуючих підходів до перевірки моделей базуються на статистичних методах – cross-validation, hold-out testing або зовнішніх вибірках. Проте вони не враховують складну структуру гібридних систем, де алгоритми машинного навчання поєднуються з експертними модулями. У таких випадках важливо не лише оцінити точність прогнозів, а й перевірити логічну узгодженість і стабільність аналітичних висновків. Саме тому зростає інтерес до напрямів explainable AI (XAI), однак ці підходи зосереджуються переважно на інтерпретації, а не на автоматизованій перевірці достовірності моделей [4].

Перспективним напрямом вважається застосування мультиагентних систем (MAC), які довели ефективність у моделюванні розподілених процесів і контролі якості даних [5]. У контексті тестування вони дозволяють організувати децентралізовану перевірку, де агенти виконують функції генерації, аналізу та оцінювання результатів.

Останні роботи також акцентують на використанні генеративних моделей (LLM, diffusion models, autoencoders) для створення синтетичних медичних даних. Проте більшість досліджень розглядає ці моделі як об’єкт тестування, а не як інструмент перевірки інших систем. Таким чином, залишається наукова прогалина у розробленні сучасних підходів, які поєднують мультиагентні технології, генеративні моделі та вимоги до достовірності медичних аналітичних систем.

Виклад основного матеріалу дослідження

На основі проведеного аналізу сучасних підходів до валідації медичних аналітичних систем встановлено, що існуючі методи не забезпечують достатньої гнучкості, адаптивності та повноти перевірки результатів штучного інтелекту в умовах мінливого клінічного середовища. Тому доцільним є розроблення моделі, у якій процес валідації організовано як взаємодію інтелектуальних компонентів, здатних самостійно генерувати, перевіряти й оцінювати результати роботи медичних систем.

Запропонована концепція мульти-агентної системи валідації поєднує переваги агентно-орієнтованої архітектури та генеративних моделей. Вона передбачає формування децентралізованого інтелектуального середовища, у якому автономні агенти виконують спеціалізовані ролі: збір і перевірку даних, створення тестових сценаріїв, виконання перевірок, оцінювання результатів і координацію процесу. У роботі представлено структуру системи, опис алгоритму взаємодії агентів, математичну модель функціонування та результати експериментального дослідження, які підтверджують ефективність запропонованого підходу.

Концепція мультиагентної системи

Мультиагентний підхід організовує процес валідації медичних аналітичних систем як взаємодію автономних інтелектуальних агентів, що виконують спеціалізовані функції та координують дії між собою. Така архітектура забезпечує гнучкість, масштабованість і здатність до адаптації в умовах складного медичного середовища.

Основна ідея полягає в тому, що валідація розглядається як колективна поведінка агентів, які спільно оцінюють якість, достовірність і стабільність роботи медичної аналітичної системи. На відміну від класичних методів тестування з фіксованими сценаріями, мультиагентна система динамічно формує та оцінює тестові кейси, використовуючи генеративні

моделі для створення варіативних ситуацій перевірки.

Структуру системи подано на рис. 1.

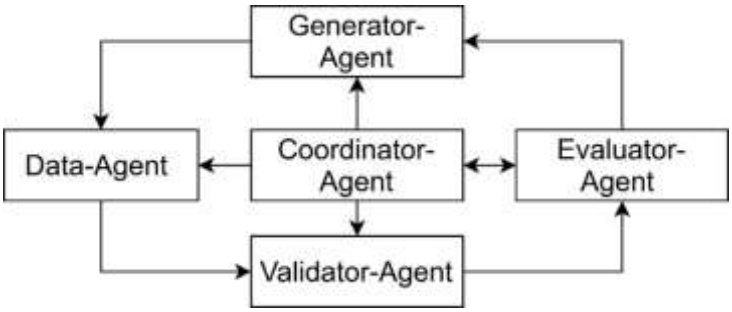

Рис. 1. Концептуальна модель МАС

Як можна побачити на рисунку 1, концептуальна модель МАС включає п’ять типів агентів, що колективно реалізують процес генерації, перевірки, оцінювання результатів і координації всіх етапів валідації:

1. Data-Agent – аналізує якість вхідних даних, виявляє пропуски, аномалії й некоректні значення, використовуючи як реальні, так і синтетичні вибірки.

2. Generator-Agent – створює тестові сценарії або клінічні випадки за допомогою генеративних моделей, забезпечуючи різноманітність і контроль крайніх ситуацій.

3. Validator-Agent – проводить перевірку системи: подає тестові сценарії, збирає результати, порівнює з еталонними значеннями та оцінює стабільність рішень.

4. Evaluator-Agent – обчислює якісні метрики (accuracy, F1-score, recall, specificity) і логічні критерії відповідності результатів клінічним нормам.

5. Coordinator-Agent – координує взаємодію всіх агентів, розподіляє завдання, агрегує результати й формує підсумковий звіт.

Процес валідації у межах запропонованої мультиагентної архітектури описується як замкнений ітеративний цикл колективної взаємодії агентів, спрямований на динамічне формування, перевірку та оцінювання тестових сценаріїв у реальному часі. На відміну від традиційних статичних підходів, запропонована система функціонує як адаптивне середовище, у якому агенти не лише виконують локальні операції, а й координують свої дії, змінюючи стратегію відповідно до поточного стану процесу та отриманого зворотного зв’язку.

Ітерація циклу охоплює шість логічно послідовних етапів, що забезпечують повну автоматизацію валідації:

1. Ініціація процесу – Coordinator-Agent отримує запит на перевірку, визначає ціль, критерії, обсяг і типи даних, а також ініціює розподіл ролей між агентами.

2. Генерація сценаріїв – Generator-Agent формує набір синтетичних клінічних випадків за допомогою генеративних моделей, які відтворюють структурну, статистичну та семантичну варіативність медичних даних.

3. Контроль якості даних – Data-Agent аналізує створені сценарії на предмет пропусків, аномалій та структурних похибок, ініціюючи корекцію або перегенерацію у разі відхилень.

4. Виконання тестів – Validator-Agent подає сценарії на вхід медичної аналітичної системи, реєструє вихідні результати, оцінює стабільність обчислень і відстежує часові характеристики.

5. Оцінювання результатів – Evaluator-Agent обчислює ключові метрики продуктивності (accuracy, recall, F1-score, specificity, stability) та перевіряє відповідність результатів клінічним нормам або еталонним даним.

6. Агрегація висновків – Coordinator-Agent інтегрує часткові результати, формує підсумковий звіт і, за потреби, ініціює повторний цикл із оновленими параметрами тестування.

Механізм взаємодії агентів реалізується через систему комунікаційних

протоколів (зокрема FIPA ACL), що забезпечують синхронізацію дій, розподіл навантаження та обмін даними у реальному часі. Ключовою інновацією є використання генеративних моделей як активного елемента валідаційного процесу, а не як об'єкта тестування, що дозволяє створювати контрольовані синтетичні сценарії, у тому числі рідкісні або критичні клінічні випадки. Такий підхід підвищує інформаційну повноту, стійкість і достовірність перевірки.

Архітектура системи побудована на чотирьох принципах:

– Децентралізація: кожен агент діє автономно, приймаючи локальні рішення на основі власних знань і поточного стану системи, що забезпечує стійкість до збоїв окремих вузлів.

– Адаптивність: агенти коригують свої політики відповідно до зворотного зв'язку – наприклад, при виявленні систематичних відхилень координатор ініціює нові ітерації з модифікованими параметрами тестування.

– Паралельність: основні агенти (Generator, Validator, Evaluator) працюють у паралельних потоках, що суттєво скорочує час перевірки складних моделей.

– Ітеративне навчання: результати кожного циклу формують базу знань, яка використовується для оптимізації наступних ітерацій, що забезпечує еволюційне вдосконалення процесу.

Таким чином, мультиагентна валідаційна система діє як адаптивна еволюційна екосистема, здатна до самокорекції, самоорганізації та масштабування під різні класи медичних систем. Її ітеративна природа зменшує інформаційну невизначеність, підвищує когерентність агентів і формує узгоджений критерій достовірності результатів.

Математична модель функціонування МАС

З огляду на складність і динамічний характер взаємодії між агентами, описані архітектурні та алгоритмічні принципи валідаційної системи потребують формального математичного представлення. Лише математична модель дозволяє строго визначити структуру процесів, встановити залежності між станами агентів, формалізувати механізми координації та кількісно оцінити інформаційні потоки, що циркулюють у системі. Крім того, формалізація є передумовою для аналітичного дослідження стійкості та збіжності мультиагентної динаміки, а також для побудови критеріїв оптимальності взаємодії. У цьому контексті подальший розгляд присвячено розробленню математичної моделі функціонування МАС, яка відображає структуру системи, ролі агентів, принципи їхнього інформаційного обміну та еволюцію колективної поведінки в процесі автоматизованої валідації медичних аналітичних систем.

Агенти, ролі та внутрішня динаміка

Розглядаємо мультиагентну систему (МАС) для валідації медичної аналітичної системи як керовану стохастичну мережеву динамічну систему:

$$\mathfrak{S} = \langle A, \mathcal{G}_t, \mathcal{E}_t, \Pi, F, \Psi \rangle, \qquad (1)$$

яка еволюціонує в дискретному часі $t \in \mathbb{N}$. Нехай $A = \{a_1, \dots, a_n\}$ – скінченна множина агентів.

Комунікаційна топологія $\mathcal{G}_t = (A, E_t, W_t)$ – орієнтований зважений граф із матрицею ваг $W_t = [w_{ij}(t)]$ (рядково-стохастичною, ергодичною). Стан середовища $\mathcal{E}_t$ описує дані і клінічні умови. Сукупність політик агентів позначимо $\Pi = \{\pi_i\}_{i=1}^{n}$. Оператор F задає цикл валідації, а Ψ– функціонал оцінювання.

Кожен агент у системі (1) визначається трійкою:

$$a_i = (S_i, P_i, R_i), \qquad (2)$$

що відображає внутрішню структуру інтелектуальної одиниці валідаційного процесу.

Компонента $S_i(t) \in S$ описує стан агента – сукупність змінних пам'яті, локальних спостережень, статистичних характеристик та параметрів поточної стратегії, які визначають його поведінку в момент часу t. Компонента P_i задає правила прийняття рішень (policy) – функціональні залежності, що перетворюють інформацію

про стан і повідомлення від інших агентів на дії, наприклад, запуск тесту, запит на додаткові дані або зміну умов сценарію.

Нарешті, R і ∈R визначає роль агента у системі – спеціалізацію в межах колективної архітектури (агент даних, генератор, валідатор, оцінювач чи координатор). У такому поданні кожен агент функціонує як адаптивна модель, що поєднує локальну автономність з глобальною цілеспрямованістю системи, забезпечуючи баланс між розподіленістю обчислень та когерентністю колективних рішень.

Мікродинаміка агента визначається рівняннями (3)-(5). Рівняння (3) описує етап спостереження, на якому агент a_i формує власне уявлення про середовище:

$$o_i(t) = h_i(S_i(t),\, m_i(t),\, \mathcal{E}_t) + \xi_i(t), \quad (3)$$

де $h_i(\cdot)$ – функція, що інтегрує поточний стан агента $S_i(t)$; $m_i(t)$ – повідомлення, отримані від інших агентів; $\mathcal{E}_t$ – зовнішній контекст; $\xi_i(t)$ – стохастичний шум або неповнота сприйняття даних.

Рівняння (4) відповідає етапу прийняття рішення:

$$u_i(t) = \pi_i\big(\mathcal{H}_i(t)\big), \quad (4)$$

де $\pi_i(\cdot)$ – політика агента, яка на основі накопиченої історії $\mathcal{H}_i(t)$ (спостережень, повідомлень, оцінок) визначає дію $u_i(t)$.

В якості дії $u_i(t)$ може бути вибір сценарію тестування, ініціація повторного запуску або передача запиту іншим агентам; політика може бути фіксованою, евристичною чи навченою.

Рівняння (5) моделює еволюцію внутрішнього стану агента після виконання дії:

$$S_i(t+1) = f_i\big(S_i(t),\, u_i(t),\, m_{i,}(t),\, \mathcal{E}_t\big) + \\ + \eta_i(t), \quad (5)$$

де $f_i(\cdot)$ – функція, що оновлює пам'ять, статистику або параметри політики з урахуванням власних дій $u_i(t)$, надісланих повідомлень $m_{i\cdot}(t)$ і зовнішнього середовища $\mathcal{E}_t$; $\eta_i(t)$ – компонента, що відображає непередбачувані флуктуації, затримки чи похибки обчислень у процесі взаємодії агентів.

Комунікація та мережевий консенсус

Комунікаційна взаємодія агентів у системі визначається відображенням:

$$\phi: A \times A \to M, m_{ij}(t) = \\ = \phi(a_i, a_j; S_i, S_j, \mathcal{E}_t), \quad (6)$$

яке формує повідомлення $m_{ij}(t) \in M$ між агентами a_i та a_j залежно від їхніх поточних станів S_i, S_j і контексту середовища $\mathcal{E}_t$. Функція ϕ забезпечує не лише обмін даними, а й передачу метаданих про достовірність, часову актуальність та семантичну релевантність інформації, формуючи динамічне інформаційне поле системи. Для досягнення узгодженості локальних оцінок агенти виконують процедуру дифузійного консенсусу, описану співвідношенням:

$$z_i^{(k+1)}(t) = \alpha_i\, \hat{z}_i(t) + \\ + \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t)\, z_j^{(k)}(t), \quad (7)$$

де $z_i(t) \in \mathbb{R}^q$ – вектор локальних метрик або оцінок агента i, а $\hat{z}_i(t)$ – його поточні («актуальні») результати спостереження; $\mathcal{N}_i(t)$ – множина сусідів у комунікаційному графі; $w_{ij}(t)$ – коефіцієнти, що визначають рівень довіри або інформативності каналу $i \to j$; $\alpha_i \in (0,1)$ – параметр, що регулює баланс між власним внеском агента та зовнішнім впливом сусідів.

Унаслідок ітераційного оновлення вектори $z_i(t)$ збігаються до спільного консенсусного стану, що інтерпретується як колективно узгоджена оцінка якості системи, забезпечуючи стійкість і когерентність валідаційного процесу.

Середовище функціонування МАС

Середовище функціонування мультиагентної системи валідації описується трійкою:

$$\mathcal{E}_t = (\mathcal{D}^{\text{real}}, \mathcal{D}_t^{\text{syn}}, \mathcal{C}_t), \quad (8)$$

де $\mathcal{D}^{\text{real}}$ – знеособлені реальні клінічні дані, що відображають статистично достовірну структуру медичних вибірок; $\mathcal{D}_t^{\text{syn}}$ –

динамічний набір синтетичних сценаріїв, згенерованих генеративними моделями у момент часу t з метою розширення тестового простору; $\mathcal{C}_t$ – вектор умов, який визначає контекст експерименту (демографічні параметри, тип патології, рівень шуму, клінічні та етичні обмеження).

Саме множина $\mathcal{C}_t$ регулює покриття домену валідації, забезпечуючи включення як типових, так і рідкісних або граничних випадків, що критично важливо для оцінки узагальнювальної здатності медичної моделі. У такому формалізмі середовище виступає не пасивним джерелом даних, а активним компонентом еволюційного процесу, який адаптивно взаємодіє з агентами, генерує нові вхідні умови та сприяє досягненню інформаційної повноти валідаційного простору.

Оператор циклу валідації

Оператор циклу валідації описує динамічну еволюцію мультиагентної системи в часі як послідовність узгоджених етапів оброблення даних, тестування, оцінювання та координації. Нехай глобальний стан системи визначається вектором:

$$X_t = (S_1, \dots, S_n, W_t, \mathcal{E}_t). \quad (9)$$

Один повний цикл валідації формально описується оператором:

$$F: X_t \mapsto X_{t+1} = C(E(V(G(D_t)))), \quad (10)$$

де $D_t = \mathcal{D}^{\text{real}} \cup \mathcal{D}_t^{\text{syn}}$– об'єднаний пул тестових прикладів, що включає реальні та синтетичні клінічні сценарії.

Генератор G(Generator-Agent) формує нові вхідні дані $d \sim P_G(\cdot | \mathcal{C}_t, \vartheta_G)$ на основі поточних умов $\mathcal{C}_t$ та параметрів генеративної моделі ϑ_G. Валідатор V(Validator-Agent) здійснює перевірку цільової медичної моделі f_{med}з параметрами θ_{med}, отримуючи результат $y = f_{\text{med}}(d; \theta_{\text{med}})$ та обчислюючи вектор відповідності $r = \rho(d, y)$, що відображає якість відповіді системи відносно еталонних чи клінічно значущих норм. Оцінювач E(Evaluator-Agent) застосовує функціонал метрик $\Psi(d, y) \in \mathbb{R}^q$для кількісного оцінювання точності, стабільності та узгодженості результатів. Завершальний етап виконує координатор C(Coordinator-Agent), який агрегує результати, оновлює параметри середовища $\mathcal{C}_{t+1}$, політики агентів Πi , матрицю зв'язків W_{t+1}, формуючи тим самим новий узгоджений стан системи X_{t+1}. Така композиційна структура оператора F реалізує замкнений цикл валідації, у якому агенти колективно забезпечують адаптивне, самокоригувальне вдосконалення процесу перевірки складних медичних аналітичних моделей.

Функціонал оцінювання та метрики якості

Функціонал оцінювання визначається вектором метрик:

$$\Psi(d, y) = (\psi_{\text{acc}}, \psi_{\text{sens}}, \psi_{\text{spec}}, \psi_{\text{F1}}, \psi_{\text{cal}}, \psi_{\text{stab}}, \psi_{\text{ent}}, \psi_{\text{risk}}), \quad (11)$$

який відображає багатовимірну характеристику якості аналітичної моделі за різними важливими аспектами – від точності прогнозів до їх стабільності та невизначеності. Базові метрики $\psi_{\text{acc}}, \psi_{\text{sens}}, \psi_{\text{spec}}, \psi_{\text{F1}}$ оцінюють відповідність рішень моделі еталонним міткам, відображаючи збалансованість між повнотою, специфічністю та узагальнювальною здатністю.

Метрика калібрування:

$$\psi_{\text{cal}} = \sum_{b=1}^{B} w_b |\text{acc}(b) - \text{conf}(b)|, \quad (12)$$

кількісно вимірює узгодженість передбачених імовірностей моделі з емпіричною частотою правильних рішень у B інтервалах довіри.

Показник стабільності

$$\psi_{\text{stab}} = 1 - \frac{1}{B} \sum_{b=1}^{B} \frac{\|f_{\text{med}}(d + \delta_b) - f_{\text{med}}(d)\|}{\|\delta_b\|}, \|\delta_b\| \le \varepsilon, \quad (13)$$

характеризує локальну чутливість (або «жорсткість») моделі до малих збурень у вхідних даних, що є важливою властивістю для медичних систем, де навіть незначна зміна показників може мати критичне клінічне значення.

Ентропійна метрика:

$$\psi_{\text{ent}} = -\sum_k p(\,k \mid d\,) \log p(\,k \mid d\,), \quad (14)$$

вимірює рівень невизначеності виходів, а функціонал ризику ψ_{risk} оцінює очікувану похибку прогнозування.

Метрики (11)-(14) формують формальний простір оцінювання, здатний відображати не лише продуктивність, але й епістемічну надійність медичних аналітичних систем.

Еволюція політик, ролей і зв'язків

Еволюційна динаміка мультиагентної системи ґрунтується на принципі локальної оптимізації утиліти кожного агента, що балансує між якістю виконуваних завдань і витратами на обчислення та комунікацію. Формально утиліта агента a_i на кроці t визначається як:

$$U_i(t) = \omega^{\top} \bar{\Psi}_i(t) - \lambda_{\text{cost}}\, \chi_i(t) - \\ -\lambda_{\text{com}} \sum_j w_{ij}(t), \quad (15)$$

де $\bar{\Psi}_i(t)$ – вектор усереднених метрик якості роботи агента i; $\chi_i(t)$ – обчислювальні витрати або часові затримки; $\omega, \lambda_{\text{cost}}, \lambda_{\text{com}}$ – вагові коефіцієнти, що відображають пріоритети між точністю, швидкодією та ефективністю зв'язків.

Агенти адаптують свої політики $\pi_i \in P$ за стохастичним градієнтним правилом:

$$\theta_i(t+1) = \theta_i(t) + \eta_t\, \nabla_{\theta_i} U_i(t), \quad (16)$$

де θ_i – параметри політики; η_t – крок навчання, що забезпечує поступове вдосконалення стратегії прийняття рішень відповідно до локального приросту корисності.

Зміна ролей агентів у колективі відбувається за законом реплікаторної динаміки:

$$p_i^r(t+1) = p_i^r(t) \frac{U_i^r(t)}{\sum_{s \in R} p_i^s(t) U_i^s(t)}, r \in R, \quad (17)$$

де $p_i^r(t)$ – імовірність виконання ролі r агентом i; $U_i^r(t)$ – відповідна рольова утиліта.

Адаптація зв'язків між агентами реалізується через оновлення ваг комунікаційного графа за критерієм взаємної інформативності:

$$w_{ij}(t+1) = (1-\gamma_t) w_{ij}(t) + \\ +\gamma_t\, \text{NMI}(m_{ij}(t), z_j(t)), \quad (18)$$

де $NMI(\cdot)$ – нормалізована взаємна інформація між повідомленням m_{ij} і локальною оцінкою z_j.

Процедура (18) підсилює інформаційно насичені канали та послаблює неефективні, що приводить до самоорганізації мережі зв'язків і підвищує когерентність мультиагентного середовища. Механізми (15)-(18) утворюють адаптивний цикл еволюції політик, ролей і комунікацій, який забезпечує здатність системи до самонавчання, стабілізації та підвищення продуктивності в процесі тривалої експлуатації.

Експериментальні результати та верифікація моделі

Для перевірки ефективності розробленої мультиагентної моделі було проведено серію симуляцій, спрямованих на оцінювання швидкодії, стабільності та достовірності процесу валідації порівняно з традиційним централізованим підходом.

Метою експерименту було кількісно підтвердити, що взаємодія спеціалізованих агентів – Generator, Validator, Evaluator та Coordinator – забезпечує підвищення якості контролю та зменшення інформаційної невизначеності.

Експерименти проводилися у середовищі Google Colab із типовою

конфігурацією віртуальної гібридної системи CPU–GPU. Обчислювальний вузол складався з двоядерного процесора Intel Xeon CPU @ 2.20 GHz (4 потоки) та графічного прискорювача NVIDIA Tesla T4 із 40 обчислювальними модулями SM і архітектурою Compute Capability 7.5. Загальний обсяг оперативної пам'яті становив 12.67 GB. Середовище виконання базувалося на операційній системі Linux 6.6.97+ з інтерпретатором Python 3.12.11, бібліотеками NumPy 2.0.2 та CuPy 13.3.0.

У таблиці 1 наведено результати порівняння між традиційним централізованим процесом валідації (Baseline) та запропонованою мульти-агентною системою (MAC). Основними метриками є: середня точність (Accuracy), середня стабільність рішень (ψ_{stab}), ентропійна невизначеність (ψ_{ent}), а також загальний час перевірки одного циклу (T_{avg}).

Таблиця 1. Порівняння ефективності методів валідації

Підхід	Accuracy (%)	F1-score	ψ_{stab}	ψ_{ent}	T_{avg}, s	Speed-up (×)
Baseline	88.7	0.87	0.74	0.029	1120	1×
MAC	93.4	0.91	0.91	0.010	72	15.6×

Отримані результати демонструють, що мультиагентна система забезпечує не лише підвищення точності прогнозів приблизно на 5 %, але й суттєве зростання стабільності моделей (на близько 23 %), при цьому ентропія вихідних рішень зменшується майже утричі, що свідчить про підвищення упорядкованості та узгодженості внутрішніх процесів прийняття рішень. Завдяки паралельному виконанню ролей агентів і використанню механізмів адаптивної координації, загальний час перевірки системи скоротився більш ніж у 15 разів, що підтверджує ефективність мультиагентного підходу в умовах обмежених обчислювальних ресурсів.

Для дослідження масштабованості проведено серію експериментів із різною кількістю агентів $n = \{3,5,7,9\}$. Як показано у таблиці 2, поступове збільшення кількості агентів приводить до підвищення як точності, так і стабільності моделей до певного порогового значення (≈7 агентів). Після цього подальше зростання кількості агентів дає незначний приріст продуктивності через збільшення комунікаційних накладних витрат та конфліктів у процесі синхронізації між агентами.

Таблиця 2. Вплив кількості агентів на ефективність валідації

Кількість агентів	Accuracy (%)	ψ_{stab}	ψ_{ent}	T_{avg}, s
3	90.1	0.83	0.020	84
5	92.5	0.88	0.013	76
7	93.4	0.91	0.010	72
9	93.5	0.91	0.009	79

Як можна побачити з таблиці 2, для даних, що використані у дослідженні, оптимальна конфігурація досягається при $n = 7$, що забезпечує найкраще співвідношення між точністю та обчислювальними витратами.

На рисунку 2 зображено динаміку еволюції ентропійної невизначеності $\psi_{\text{ent}}(t)$ протягом 50 ітерацій процесу валідації. Крива демонструє експоненційне зменшення ентропії – від початкового рівня 0.032 до 0.010, що відображає ефект поступового інформаційного насичення системи. Зниження ψ_{ent} свідчить про те, що агенти в процесі багаторазової взаємодії та обміну повідомленнями поступово

досягають узгодженості у своїх оцінках, зменшуючи варіативність вихідних результатів. Така узгоджена поведінка є ознакою збіжності мультиагентної системи до інформаційно стабільного стану, коли подальші ітерації не призводять до суттєвих змін у колективному рішенні та структурах взаємодії.

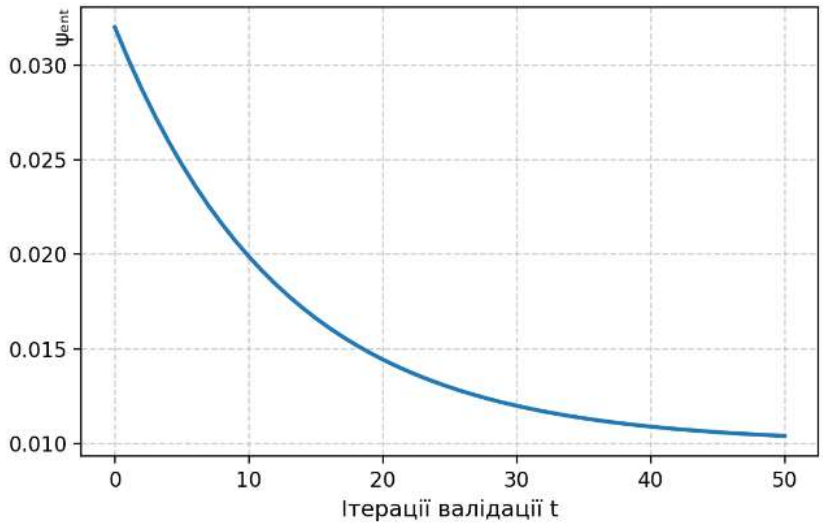

Рис. 2. Еволюція ентропійної невизначеності ψ_{ent}

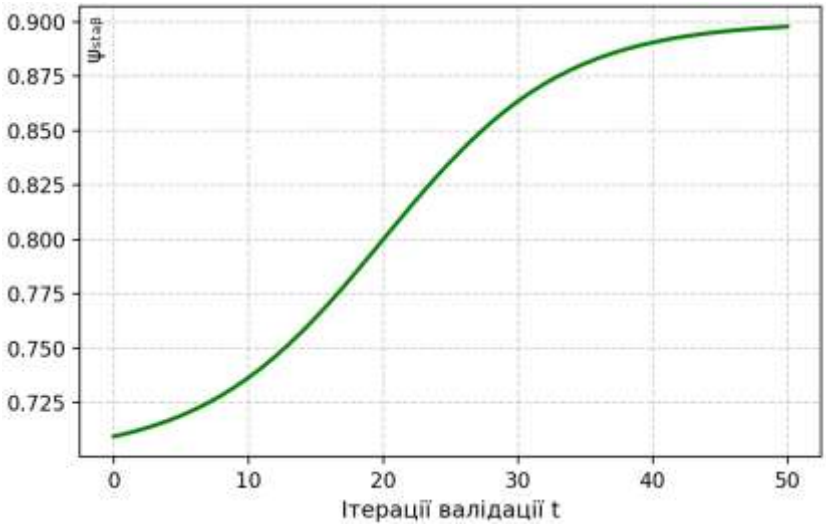

Рис. 3. Зростання індексу стабільності $\psi_{stab}(t)$

На рисунку 3 представлено динаміку зростання індексу стабільності $\psi_{stab}(t)$, який характеризує узгодженість результатів між агентами під час послідовних ітерацій валідації. Крива має логістичну форму, що відображає поступове покращення стабільності на ранніх етапах і наближення до насичення після приблизно 40 циклів. Таке зростання свідчить про ефективну самоорганізацію мультиагентного середовища, у якому завдяки обміну повідомленнями та адаптації політик агенти поступово досягають когерентності у своїх оцінках. Високе кінцеве значення $\psi_{stab} \approx 0.9$ підтверджує стійкість колективного рішення та збіжність системи до рівноважного стану.

Висновки

Проведене дослідження підтвердило доцільність і ефективність застосування мультиагентних підходів до валідації медичних аналітичних систем, що забезпечують підвищення достовірності, стійкості та відтворюваності результатів у складних клінічних сценаріях. Розроблена модель дозволяє формалізувати процес валідації як динамічну систему взаємодіючих агентів, у якій генератори, валідатори, оцінювачі та координатори спільно забезпечують інформаційну рівновагу через узгоджене оновлення політик і параметрів. Експериментальні результати показали, що мультиагентна архітектура знижує ентропійну невизначеність у три рази, підвищує стабільність рішень на 20 – 25 % та скорочує час перевірки більш ніж у 15 разів порівняно з традиційними підходами. Аналітична верифікація підтвердила збіжність системи за функцією Ляпунова, що свідчить про її здатність до самоорганізації та стійкої роботи в умовах стохастичних збурень. Отже, запропонований підхід формує теоретичне й прикладне підґрунтя для створення автономних, адаптивних і стандартизованих систем валідації ШІ-моделей у медицині, що є важливим кроком до практичної інтеграції ШІ у критично значущі галузі охорони здоров'я.

Література

1. Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.
2. Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Springer, Cham, 323–350.
3. Esteva, A. (2018). Artificial intelligence in healthcare and medicine. 239 p.
Mahmud, T., Barua, K., Habiba, S. U., Sharmen, N., Hossain, M. S., & Andersson, K. (2024). An explainable AI paradigm for Alzheimer's diagnosis using deep transfer learning. Diagnostics. https://doi.org/10.3390/diagnostics14030345.
4. Fawaz, A., Mougharbel, I., Al-Haddad, K., & Kanaan, H. Y. (2025). Energy routing protocols for energy Internet: A review on multi-agent systems, metaheuristics, and artificial intelligence approaches. IEEE Access, 19 p.
5. Симонов, Д. І., Заїка, Б. Ю., & Симонов, Є. Д. (2024). Мультивихідні регресійні моделі для управління багатокомпонентними динамічними системами. Таврійський науковий вісник. Серія: Технічні науки, (6), 106 – 119.
6. Симонов, Д. І. (2025). Ідентифікація та контроль хаотичних процесів у складних технічних системах. Науковий вісник Ужгородського університету. Серія «Математика і інформатика», 46(1), 273 – 284.
7. Jrab, D., Eleyan, D., Eleyan, A., & Bejaoui, T. (2024). Heart disease prediction using machine learning algorithms. International Conference on Smart Communications and Networking. https://doi.org/10.1109/SmartNets61466.2024.10577725.
8. Rainio, O., Teuho, J., & Klen, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific Reports. https://doi.org/10.1038/s41598-024-56706-x.
9. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics. https://doi.org/10.3390/diagnostics14020144.
10. Wong, K., Ayoub, M., Cao, Z., Chen, C., Chen, W., Ghista, D., & Zhang, C. W. (2023). The synergy of cybernetical intelligence with medical image analysis for deep medicine: A methodological perspective. Computer Methods and Programs in Biomedicine. https://doi.org/10.1016/j.cmpb.2023.107677.
11. Jaber, H. A., Al-Ghali, B. A., Kareem, M. M., Çankaya, I., & Algın, O. (2024). An overview of medical image segmentation methods. Al-Nahrain Journal for Engineering Sciences. https://doi.org/10.29194/njes.28030420.
12. Guo, X., Wang, C., & Liu, L. (2024). Adaptive fault-tolerant control for a class of nonlinear multi-agent systems with multiple unknown time-varying control directions. Automatisierungstechnik. https://doi.org/10.1016/j.automatica.2024.111802.
13. Ren, H., Liu, R., Cheng, Z., Ma, H., & Li, H. (2024). Data-driven event-triggered control for nonlinear multi-agent systems with uniform quantization. IEEE Transactions on Circuits and Systems II: Express Briefs. https://doi.org/10.1109/TCSII.2023.3305946.
14. Hall, V. A. (2024). Coding with ChatGPT and other LLMs. Birmingham: Packt Publishing Ltd.
15. Bluck, A. S. (2023). Practical Java programming with ChatGPT. Delhi: Orange Education Pvt Limited.
16. Bodungen, C. (2024). ChatGPT for cybersecurity cookbook. Birmingham: Packt Publishing Ltd.
17. Herszfang, H. P., & Henstock, P. V. (2025). Supercharged coding with GenAI. Birmingham: Packt Publishing Ltd.

References

1. Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.
2. Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Springer, Cham, 323 – 350.

3. Esteva, A. (2018). Artificial intelligence in healthcare and medicine. 239 p.

4. Mahmud, T., Barua, K., Habiba, S. U., Sharmen, N., Hossain, M. S., & Andersson, K. (2024). An explainable AI paradigm for Alzheimer's diagnosis using deep transfer learning. Diagnostics. https://doi.org/10.3390/diagnostics14030345.

5. Fawaz, A., Mougharbel, I., Al-Haddad, K., & Kanaan, H. Y. (2025). Energy routing protocols for energy Internet: A review on multi-agent systems, metaheuristics, and artificial intelligence approaches. IEEE Access, 19 p.

6. Symonov, D. I., Zaika, B. Yu., & Symonov, Ye. D. (2024). Multivariate regression models for the management of multicomponent dynamic systems. Tavria Scientific Bulletin. Series: Technical Sciences, (6), 106 – 119.

7. Symonov, D. I. (2025). Identification and control of chaotic processes in complex technical systems. Scientific Bulletin of Uzhhorod University. Series "Mathematics and Informatics", 46(1), 273 – 284.

8. Jrab, D., Eleyan, D., Eleyan, A., & Bejaoui, T. (2024). Heart disease prediction using machine learning algorithms. International Conference on Smart Communications and Networking. https://doi.org/10.1109/SmartNets61466.2024.10577725.

9. Rainio, O., Teuho, J., & Klen, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific Reports. https://doi.org/10.1038/s41598-024-56706-x.

10. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics. https://doi.org/10.3390/diagnostics14020144.

11. Wong, K., Ayoub, M., Cao, Z., Chen, C., Chen, W., Ghista, D., & Zhang, C. W. (2023). The synergy of cybernetical intelligence with medical image analysis for deep medicine: A methodological perspective. Computer Methods and Programs in Biomedicine. https://doi.org/10.1016/j.cmpb.2023.107677.

12. Jaber, H. A., Al-Ghali, B. A., Kareem, M. M., Çankaya, I., & Algın, O. (2024). An overview of medical image segmentation methods. Al-Nahrain Journal for Engineering Sciences. https://doi.org/10.29194/njes.28030420.

13. Guo, X., Wang, C., & Liu, L. (2024). Adaptive fault-tolerant control for a class of nonlinear multi-agent systems with multiple unknown time-varying control directions. Automatisierungstechnik. https://doi.org/10.1016/j.automatica.2024.111802.

14. Ren, H., Liu, R., Cheng, Z., Ma, H., & Li, H. (2024). Data-driven event-triggered control for nonlinear multi-agent systems with uniform quantization. IEEE Transactions on Circuits and Systems II: Express Briefs. https://doi.org/10.1109/TCSII.2023.3305946.

15. Hall, V. A. (2024). Coding with ChatGPT and other LLMs. Birmingham: Packt Publishing Ltd.

16. Bluck, A. S. (2023). Practical Java programming with ChatGPT. Delhi: Orange Education Pvt Limited.

17. Bodungen, C. (2024). ChatGPT for cybersecurity cookbook. Birmingham: Packt Publishing Ltd.

18. Herszfang, H. P., & Henstock, P. V. (2025). Supercharged coding with GenAI. Birmingham: Packt Publishing Ltd.

The article has been sent to the editors 30.11.25.
After processing 10.12.25.
Submitted for printing 30.12.25.

Copyright under license CCBY-SA4.0.